МАТЕМАТИЧКА СТАТИСТИКА - Решење колоквијума

1. Нека је $\mathbf{X}$ узорак из фамилије расподела $\mathcal{N}(\sigma,\sigma), \sigma>0$

а) Показати да је $\sum_{i=1}^n X_i^2$ минимална довољна статистика за параметар σ.

б) Наћи функцију $g(\sigma)$ такву да је $\sum_{i=1}^n X_i^2$ јединствена најбоља оцена за параметар $g(\sigma)$.

в) Одредити доњу границу дисперзија Рао-Крамера за $g(\sigma)$. Да ли је ова оцена достиже?

Напомена: Ако је X из фамилије нормалних расподела $\mathcal{N}(m,\sigma^2)$, $EX^4 = m^4+6m^2\sigma^2+3\sigma^4$.

Решење:

а)

$$f(x;\sigma)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\sigma)^2}{2\sigma}}=e^{-\frac{x^2}{2\sigma}+x-\frac{\sigma}{2}+\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln\sigma}$$

Наведена расподела задовољава услове регуларности, па припада регуларној експоненцијалној фамилији расподела. Према томе је $\sum_{i=1}^n X_i^2$ комплетна довољна статистика за σ. Она је и минимална јер је једнодимензиона.

б) Нађимо очекивање дате статистике.

$$E\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)=E(X^2)=\sigma+\sigma^2$$

Други начин (без експлицитног коришћења момената нормалне расподеле): Како је $T(X)$ статистика која фигурише у запису једнопараметарске експоненцијалне фамилије расподела њено математичко очекивање је

$$\frac{d'(\sigma)}{c'(\sigma)}=\frac{(\frac{\sigma}{2}+\frac{1}{2}\ln\sigma)'}{(-\frac{1}{2\sigma})'}=\sigma+\sigma^2.$$

Пошто је $\sum_{i=1}^n X_i^2$ комплетна довољна статистика за σ, то је и $T=\frac{1}{n}\sum_{i=1}^n X_i^2$ као њена 1-1 функција. T је непристрасна оцена за $g(\sigma)=\sigma+\sigma^2$, па је према теореми Леман-Шефеа јединствена најбоља оцена за $g(\sigma)$.

Тражена функција је $g(\sigma)=\sigma+\sigma^2$

в) Пошто фамилија припада регуларној експоненцијалној фамилији расподела, она је регуларна и у смислу Рао-Крамера.

$\ln f(x;\sigma)=-\frac{x^2}{2\sigma}+x-\frac{1}{2}\sigma-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln\sigma$

$\frac{d\ln f(x;\sigma)}{d\sigma}=\frac{x^2}{2\sigma^2}-\frac{1}{2}-\frac{1}{2\sigma}$

$\frac{d^2\ln f(x;\sigma)}{d\sigma^2}=-\frac{x^2}{\sigma^3}+\frac{1}{2\sigma^2}$

$I(\sigma)=-E(\frac{d^2\ln f(X;\sigma)}{d\sigma^2})=\frac{1}{\sigma^3}EX^2-\frac{1}{2\sigma^2}=\frac{1}{\sigma^3}(\sigma+\sigma^2)-\frac{1}{2\sigma^2}=\frac{1+2\sigma}{2\sigma^2}$

Доња граница дисперзија Рао-Крамера за $g(\sigma)$ је: $G=\frac{(g'(\sigma))^2}{nI(\sigma)}=\frac{2\sigma^2(1+2\sigma)}{n}$

$D(\frac{1}{n}\sum_{i=1}^n X_i^2)=\frac{1}{n}D(X^2)=\frac{1}{n}(E(X^4)-(EX^2)^2)=\frac{1}{n}(\sigma^4+6\sigma^3+3\sigma^2-(\sigma+\sigma^2)^2)=\frac{1}{n}(4\sigma^3+2\sigma^2)$

Пошто је $D(\frac{1}{n}\sum_{i=1}^n X_i^2)=G$ оцена достиже доњу границу Рао-Крамера.

Други начин (без рачунања границе и дисперзије оцене): Граница Рао-Крамера се достиже само ако је узорачка расподела из једнопараметарске експоненцијалне фамилије расподела, а оцена чија дисперзија достиже границу је управо довољна статистика која фигурише у запису те расподеле. Како су сви ови услови испуњени, дисперзија оцене $\frac{1}{n}\sum_{i=1}^n X_i^2$ достиже границу Рао-Крамера.

2. Нека је $\mathbf{X}$ узорак из фамилије логистичких расподела с густином

$$f(x;\theta)=\frac{e^{-\frac{x}{\theta}}}{\theta(1+e^{-\frac{x}{\theta}})^2},\ x\in\mathbb{R},\theta>0$$

.

а) Показати да је $EX_i=0$.

б) Ако је $DX_i=\frac{\theta^2\pi^2}{3}$, одредити оцену T_n параметра θ методом момената.

в) Доказати да је T_n постојана оцена.

Решење:

а)

$$\begin{aligned}
E(X_i) &= \int_{-\infty}^{\infty} x\frac{e^{-\frac{x}{\theta}}}{\theta(1+e^{-\frac{x}{\theta}})^2}dx = \left|\begin{array}{c} 1+e^{-\frac{x}{\theta}}=t \\ -\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=dt \end{array}\right| \\
&= -\theta\int_1^{\infty}\frac{\ln(t-1)}{t^2}dt = \left|\begin{array}{c} u=\ln(t-1), du=\frac{dx}{t-1} \\ dv=\frac{1}{t^2}, v=-\frac{1}{t} \end{array}\right| \\
&= -\theta\left(-\frac{1}{t}\ln(t-1)|_1^{\infty}-\int_1^{\infty}\frac{1}{t(t-1)}dt\right) \\
&= -\theta\left(-\frac{1}{t}\ln(t-1)|_1^{\infty}-\int_1^{\infty}\frac{dt}{t}+\int_1^{\infty}\frac{dt}{t-1}\right) \\
&= -\theta\ln\left(\frac{(t-1)^{1-\frac{1}{t}}}{t}\right)|_1^{\infty}=0,
\end{aligned}$$

Други начин је да се покаже да је функција симетрична око 0 и да интеграл (апсолутно) конвергија. Симетрија важи јер је

$$f(-x;\theta)=\frac{e^{\frac{x}{\theta}}}{\theta(1+e^{\frac{x}{\theta}})^2}=\frac{e^{\frac{x}{\theta}}}{\theta e^{2\frac{x}{\theta}}(e^{-\frac{x}{\theta}}+1)^2}=f(x;\theta)$$

б) $T_n=\frac{\sqrt{3}}{\pi}\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}$ или $T_n=\frac{\sqrt{3}}{\pi}\overline{S}_n$

в) Ако се користи прва од две наведене оцене, из закона великих бројева следи

$$\frac{1}{n}\sum_{i=1}^n X_i^2\xrightarrow{P}EX^2=\frac{\theta^2\pi^2}{3}.$$

Из теореме о непрекидном пресликавању (примењене на функцију $g(x)=\frac{\pi}{\sqrt{3}}\sqrt{x}$) имамо

$$T_n=g\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)\xrightarrow{P}g\left(\frac{\theta^2\pi^2}{3}\right)=\theta$$

Дакле T_n је постојана оцена за θ.

Ако се користи друга оцена, онда из закона великих бројева имамо $\frac{1}{n}\sum_{i=1}^n X_i^2\xrightarrow{P}\frac{\theta^2\pi^2}{3}$ и $\overline{X}_n\xrightarrow{P}0$, па применом теореме Слуцког имамо

$$\overline{S}_n=\frac{1}{n}\sum_{i=1}^n X_i^2-\overline{X}_n^2\xrightarrow{P}\frac{\theta^2\pi^2}{3}.$$

(Коришћено је да конвергенција у расподели повлачи конвергенцију у вероватноћи, као и да ако низ конвергира у расподели ка константи онда конвергира и у вероватноћи ка тој константи). Даље, основу теореме о непрекидном пресликавању, примењене као у случају прве оцене, имамо тражену постојаност.